

MACHINE LEARNING

MCQ.

1. Which of the following methods do we use to find the best fit line for data in Linear Regression? A) Least Square Error B) Maximum Likelihood C) Logarithmic Loss D) Both A and B

Ans. A) Least square error.

2. Which of the following statement is true about outliers in linear regression? A) Linear regression is sensitive to outliers B) linear regression is not sensitive to outliers C) Can't say D) none of these

Ans. A) Linear regression is sensitive to outliers

3. A line falls from left to right if a slope is _____? A) Positive B) Negative C) Zero D) Undefined

Ans. B) Negative

4. Which of the following will have symmetric relation between dependent variable and independent variable? A) Regression B) Correlation C) Both of them D) None of these

Ans. B) Correlation

5. Which of the following is the reason for over fitting condition? A) High bias and high variance B) Low bias and low variance C) Low bias and high variance D) none of these

Ans. C) Low bias and high variance

6. If output involves label then that model is called as: A) Descriptive model B) Predictive modal C) Reinforcement learning D) All of the above

Ans. B) Predictive modal

7. Lasso and Ridge regression techniques belong to _____? A) Cross validation B) Removing outliers C) SMOTE D) Regularization

Ans. D) Regularization

8. To overcome with imbalance dataset which technique can be used? A) Cross validation B) Regularization C) Kernel D) SMOTE

Ans. D) SMOTE

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph? A) TPR and FPR B) Sensitivity and precision C) Sensitivity and Specificity D) Recall and precision

Ans. A) TPR and FPR

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less. A) True B) False

Ans. B) False

11. Pick the feature extraction from below: A) Construction bag of words from a email B) Apply PCA to project high dimensional data C) Removing stop words D) Forward selection

Ans. A) Construction bag of words from a email

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression? A) We don't have to choose the learning rate. B) It becomes slow when number of features is very large. C) We need to iterate. D) It does not make use of dependent variable.

Ans. D) It does not make use of dependent variable.

Subjective answer type questions

13. Explain the term regularization?

Ans. Regularization in machine learning is a technique used to prevent overfitting by adding a penalty term to the loss function during model training. It helps to control the complexity of the model and improve its generalization performance on unseen data. Common regularization methods include L1 regularization (Lasso) and L2 regularization (Ridge), which penalize large coefficients in the model. Regularization encourages simpler models that are less prone to overfitting.

14. Which particular algorithms are used for regularization?

Ans. Regularization techniques can be applied to various machine learning algorithms, but they are most commonly associated with linear regression and logistic regression. In these cases, regularization helps to control the complexity of the models and prevent overfitting.

Specifically, regularization techniques such as L1 regularization (Lasso) and L2 regularization (Ridge) are commonly used with linear regression and logistic regression. These techniques penalize large coefficients in the models, encouraging simpler and more generalizable solutions.

However, regularization can also be applied to other algorithms, such as support vector machines, neural networks, and decision trees, to improve their generalization performance and prevent overfitting. For example, regularization parameters can be added to the loss function of neural networks to penalize large weights and biases.

Overall, while regularization is commonly associated with linear and logistic regression, it is a widely used technique that can be applied to various machine learning algorithms to improve their performance.

15. Explain the term error present in linear regression equation?

Ans. In machine learning, the term error in the linear regression equation refers to the difference between the predicted values generated by the model and the actual observed values in the dataset. It represents how much the model's predictions deviate from the true values. The goal of linear regression is to minimize these errors across all data points by optimizing a chosen loss function, such as mean squared error or mean absolute error. Analyzing these errors helps assess the model's performance and identify areas for improvement.

